CDC

MANUAL
6

SELF-STUDY COURSE 3030-G

Principles of Epidemiology

Investigation of Disease Outbreaks
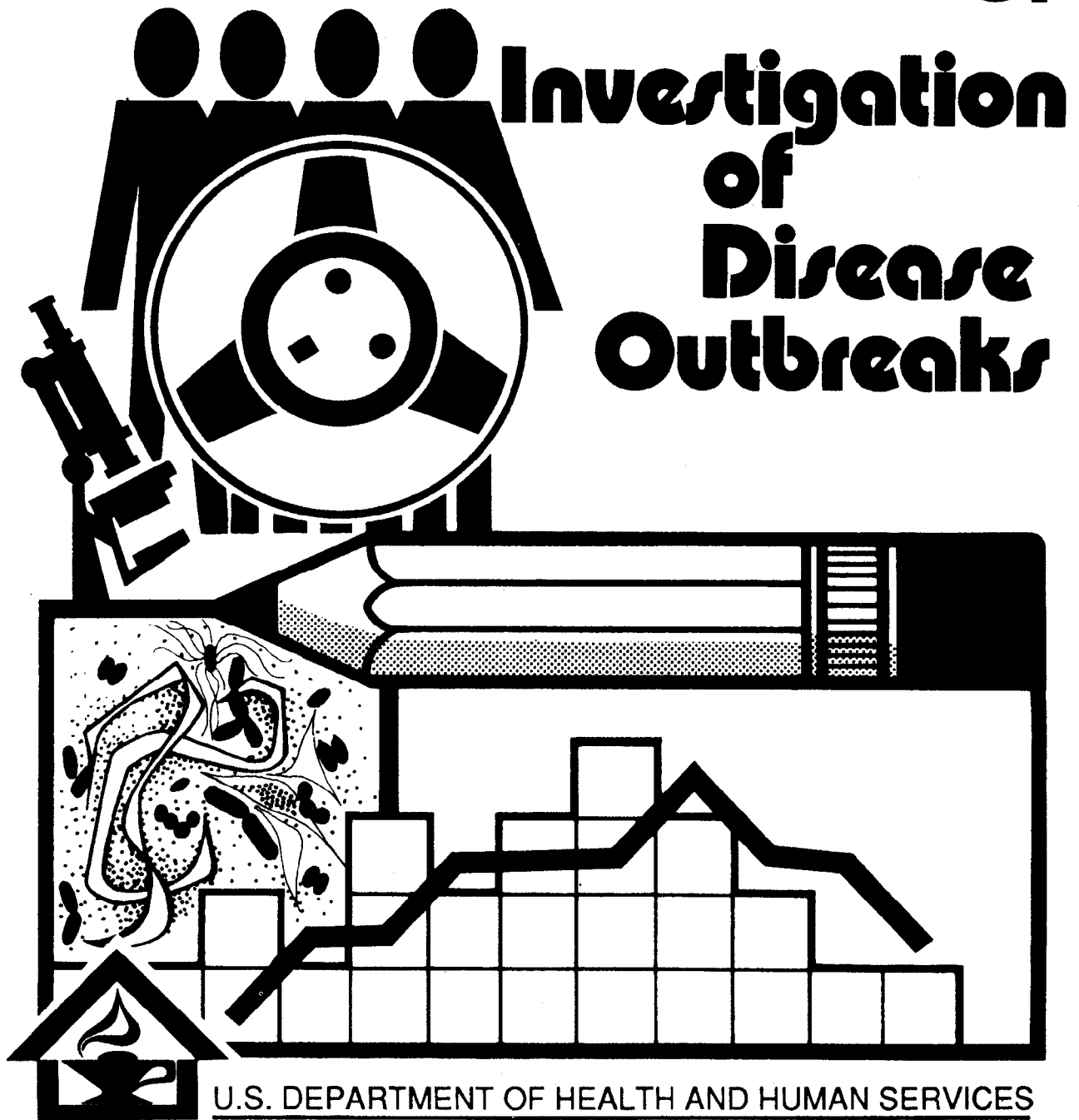
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

SELF-STUDY

TABLE OF CONTENTS

INVESTIGATION OF DISEASE OUTBREAKS

PRINCIPLES OF EPIDEMIOLOGY

Self-Study Course 3030-G

INVESTIGATION OF DISEASE OUTBREAKS

INTRODUCTION

The primary objective of an epidemic or disease outbreak investigation
usually is to identify ways of preventing the further transmission of
the disease-causing agent. In order to achieve the primary objective it
is necessary to achieve various subobjectives. These are to:

1. Establish or verify the diagnosis of reported cases, and identify
    the specific etiologic agent responsible.
2. Confirm that an outbreak or an epidemic exists.
3. Describe the cases in the epidemic or outbreak according to the
    variables of time, place, and person.
4. Identify the source of the agent and its mode of transmission,
    including the specific vehicles, vectors and routes that may
    have been involved.
5. Identify susceptible populations that are at an increased risk
    of exposure to the agent.

The objectives and methods of epidemic or disease outbreak investigations,
except for subobjective #2, "Confirm that an epidemic exists", are also
relevant to the study of endemic disease situations.

The sequence of these subobjectives indicates the sequence in which the
logic proceeds in an epidemiologic investigation, but it is not necessarily
the sequence in which the investigation itself is conducted. In practice,
several steps of the investigation may be in progress simultaneously. The
sequence shown would more likely be followed in those instances in which
an investigation was conducted some time after an epidemic had ended. An
investigation consists of repetitions of the following process until the
previously stated objectives have been met:

1. Assemble and organize the available information so that it
    can be analyzed.
2. Draw conclusions from that information.
3. Assemble your conclusions into a hypothesis.
4. Identify the specific additional information needed to test
    the hypothesis.
5. Obtain that information and test the hypothesis.
6. Start over at step #1 if appropriate.

For example, if nine cases of salmonellosis were reported to a health
department during one particular week, one could conclude that this is
an unusual event and hypothesize that an outbreak of salmonellosis is
in progress. Additional information would have to be obtained to test
this hypothesis:

1

1. Signs and symptoms of the cases--and appropriate specimens for laboratory analysis--to confirm the diagnosis and to identify the specific agent responsible.
2. Date of onset of the illness in each case--to confirm that the cases are current and that the apparent clustering of cases in time is not an artifact of the reporting system itself.
3. The usual incidence of salmonellosis in the community--so that a judgement can be made as to whether the current incidence represents an excess (assuming that an excess of cases is the criterion for conducting an investigation).

Once obtained, this information must be organized and interpreted. If, on the basis of the new data, you conclude the cases are current, they do have salmonellosis, and the current incidence does represent an excess, then you have successfully tested your hypothesis and found it to be valid (and met subobjectives 1 and 2).

Successive cycles of the above 6-step process would focus on identifying (1) the source of infection for the cases, (2) the mode and vehicle or vector by which the agent was transmitted, and (3) the circumstances leading to the presence of the agent at the source. Knowing these things, one then can identify high-risk populations and implement appropriate control measures.

It may be that not all persons involved in an epidemiologic investigation are in a position to implement the appropriate controls, or even to recommend them. This could result from an inadequate number of cases of a disease locally upon which to form any meaningful conclusions; from a lack of knowledge of appropriate control measures; from a failure to obtain the information necessary to meet the objectives of the investigation; or from the fact that the individual's role in the investigation does not include that responsibility. Nevertheless, accurate, complete, epidemiologic information is essential for the identification, recommendation and implementation of effective and efficient control measures and for their subsequent evaluation. Many epidemiologists feel very strongly that no outbreak or epidemic investigation is complete or successful until such measures have been implemented.

Step 1: Establishing or Verifying the Diagnosis

In order to develop accurate case counts for later analyses, it is essential to verify the diagnosis of reported cases associated with a suspected epidemic. The reasons for the great importance attached to this step are that:

1. Medicine being an inexact science, diseases can be misdiagnosed.
2. You may have been informed of the occurrence not of a case, but of a suspect or of a person or persons having a particular syndrome.
3. Information from non-cases (i.e., reported cases whose diagnosis could not be confirmed) must be excluded from the case information used to confirm the presence or absence of an epidemic.

A diagnosis made upon clinical grounds only can be subject to considerable error. Signs and symptoms of many diseases are not so specific that a diagnosis can be definitely established on these grounds alone. Another complication is that many ill persons do not manifest the typical syndrome of their illness. For some illnesses a clinical diagnosis is not sufficient—one must also know the specific serotype (or other classification) of the agent causing the illness. The reason is that many serotypes of a species of an infectious agent might be simultaneously present in the community.

Whenever possible, then, laboratory tests should be conducted to confirm the diagnosis. For those diseases for which no specific diagnostic laboratory tests are available, the laboratory can sometimes be used to exclude cases having a disease which produces a syndrome similar to that of the disease suspected. Since some laboratory confirmations may require weeks, with the implementation of control measures delayed correspondingly, an interim diagnosis often is made on the basis of the clinical and epidemiologic characteristics of the cases.

With or without supportive diagnostic laboratory tests, specific criteria—in terms of signs and symptoms (Appendix A)—must be developed to define a case. In Appendix B are presented several examples of case-investigation forms. A key feature of each is a section for the signs and symptoms of the reported case. When such a form is completely and properly filled out, the signs and symptoms, and the results of any laboratory work done, can be reviewed to determine whether the person has the suspected disease. The review of these data must be done using objective criteria of a case. Those persons who do not meet the criteria can be dropped—at least temporarily—from consideration. Those persons who meet the criteria for a case could be simply identified as a case or they could be further categorized as to definite, probable, or suspect cases. This categorization would be done on the basis of whether or not the clinical diagnosis was supported by laboratory tests, and on the number, nature and severity of the signs and symptoms.

It is often useful, especially when a diagnosis has not yet been established or confirmed, to calculate the frequency distribution of the signs and symptoms of the cases. This is done by first listing each of the various signs and symptoms that the cases reported. Second, count the number of cases that had each specific sign or symptom. Then calculate the % of cases that had each. For easier interpretation of the results, the signs and symptoms should be arranged in descending order of frequency (see table 1). Additional examples are shown in Appendix C.

Table 1

The Frequency of Signs and Symptoms in Cases of Shigellosis in Three Outbreaks: (1) Seattle, Washington, July 1976, (13 Cases); (2) Wayne County, Pennsylvania, August 9-11, 1974, (Approximately 480 Cases); and (3) Tourists to Jamaica, November 7-15, 1970 (63 Cases).

Signs and Symptoms	% of Cases Having Sign or Symptom		
	(1)	(2)	(3)
1. Diarrhea	100	94	100
2. Vomiting	-	73	13
3. Abdominal pain or cramps	46	70	16
4. Headache	46	18	-
5. Fever	46	-	27
6. Chills	39	-	-
7. Nausea	-	-	35
8. Dizziness	-	25	-
9. Prostration	-	15	-
10. Myalgia	8	-	-
11. Bloody stools	8	-	5

Sources: (1) HEW, PHS, CDC, Morbidity and Mortality Weekly Report, October 1, 1976.
(2) HEW, PHS, CDC, Morbidity and Mortality Weekly Report, September 20, 1975.
(3) HEW, PHS, CDC, Morbidity and Mortality Weekly Report, Vol. 20, p. 35; 1971.

The objective of establishing or verifying the diagnosis of a reported case of a disease, then, will be considered to have been reached when your criteria for a case of that disease have been matched against the suspect's signs, symptoms and laboratory test results, and a conclusion is reached as to whether the reported case is or is not a true case.

Step 2: Confirming the Existence of an Epidemic or of a Disease Outbreak

To decide whether an epidemic, or a disease outbreak, exists, a comparison is made between the current incidence and the usual incidence of cases among the population determined to be at risk of infection. If the current incidence is markedly in excess of the usual incidence, an epidemic is usually considered to exist. Small differences between usual and current incidences can cause uncertainty, so the investigator should always be on the lookout for newly developing cases that can confirm a suspected epidemic.

Clearly the term "excess" is imprecise. This usually is not a problem in the recognition of large common source outbreaks, but it does pose a problem to the early detection of outbreaks of either propagated source or of vectorborne diseases.

When an epidemic is only suspected, the population at risk frequently is not known with any high degree of accuracy. Therefore, at least initially, the population at risk is commonly assumed, for the purpose of identifying the usual and the current incidence, to be the entire population of the particular geographic area or institution in which the disease is occurring. If the suspected outbreak is known or thought to be occurring in a highly restricted population--a school, nursing home, day care center or civic or social group--then the available information for the current and the usual incidence in that group should be used to determine whether an outbreak exists.

Current Incidence

When an epidemic is suspected, an initial count of the current cases (those persons whose infection or poisoning occurred during the period of the suspected epidemic) is made in order to confirm the existence of an "excess" frequency of new cases. At the time of the initial case count there may not be sufficient information available about each case to confirm the diagnosis. In this situation the appropriate thing to do is to make sure that:

1. whatever needs to be done to establish a diagnosis is being done--for each case; and,
2. the suspected cases included in the initial count at least have some specified signs and symptoms in common.

While contacting the various sources of data to obtain, if necessary, more details about the characteristics of the reported cases, the investigator should also encourage the immediate reporting of cases which subsequently come to their attention. That is, expand the local surveillance immediately-- to all possible data sources.

Morbidity reports received by the health department are a readily available source of information for the case count. In addition to health department records, other additional sources such as doctors, hospitals and clinics, and laboratories, are often important.

Contacts with doctors sometimes disclose cases which were diagnosed but not reported, as well as possible cases for which a diagnosis has not as yet been established. Hospitals and clinics may provide clinical and laboratory information concerning those cases which were hospitalized. Laboratories are an important source of information concerning preliminary or final test results conducted on possible cases. They should be encouraged to report suspects' diagnostic test results as rapidly as possible.

Known cases and their associates are important sources of information about additional undiagnosed or unreported cases. Cases which are interviewed may provide clues to the existence of sub-clinical or clinical cases among immediate family members, relatives, and extrafamilial associates. The interview may also lead to the source of infection as well as associates developing the disease for whom the known case might have served as the source.

The existence of an epidemic is generally considered to have been confirmed if the incidence of a particular disease during a specified period of time is demonstrably in excess of the usual incidence of that disease in that population during past periods of comparable length. The definition of "excess" is left to each health jurisdiction, and different jurisdictions may view a given situation differently from each other.

Step 3: Characterizing the Epidemic

As indicated earlier, the epidemic must be characterized according to the variables of time, place, and person. The characterization must be done in such a manner that a hypothesis can be developed regarding the source, mode of transmission, and duration of the epidemic. To formulate the necessary hypotheses, the initial case information collected must be organized in such a manner as to permit answers to the following kinds of questions:

a. Regarding time:
1. What is the exact period of the outbreak?
2. Given the diagnosis, what is the probable period of exposure?
3. Is the outbreak most likely common source or propagated or both?

b. Regarding place:
1. What is the most significant geographic distribution of cases (by place of residence? Work? Some other place?)
2. What are the attack rates?

c. Regarding the characteristics of the persons (cases) involved?
1. What were the age- and sex-specific attack rates?
2. What age- and sex-groups were at the highest and lowest risk of illness?
3. In what other ways do the characteristics of the cases differ significantly from those of the general population?

Time

Variation in the frequency of occurrence of cases of a disease in a population over time is commonly referred to as that disease's temporal disease pattern. There are three basic spans of time used to describe temporal disease patterns: the epidemic period, which is of variable length, depending upon the duration of the particular epidemic; a 12-month period, to identify seasonal variation; and an indefinitely long period of years, to identify long-term trends. Seasonal variations and long-term trends of disease incidence are important considerations in confirming or rejecting the existence of an epidemic during a current period and in predicting future epidemic periods.

The rest of this discussion of time as an epidemiologic variable will focus on the construction and use of an epidemic curve. An epidemic curve is constructed mainly to:

a. determine whether the source of infection probably was common or propagated or both; and
b. identify the probable time of exposure of the cases to the source of the infection.

7

For the purpose of this discussion, an epidemic curve is defined as a
graph in which the cases of a disease which occurred during an epidemic
period are graphed according to the time of onset of illness in the cases
(as in Figure 1).

Figure 1

Date of Onset of Illness of 37 cases of Rubella, Sun City, June 21-29
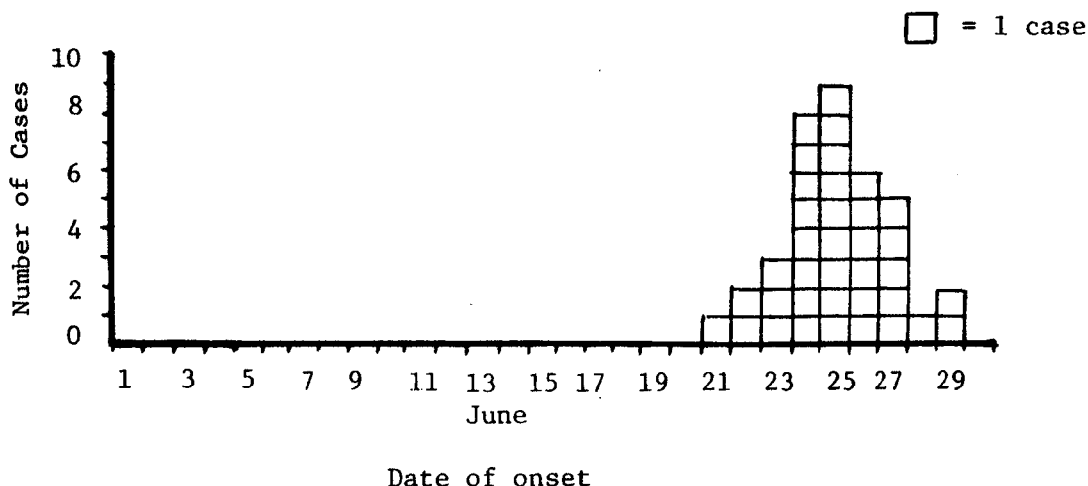
Date of onset

To draw an epidemic curve, the date of onset of illness of the cases must
be obtained. With certain diseases having a very short incubation or
latency period, the hour must be obtained for each case. Next select an
interval of time by which the cases will be graphed. Appropriate time
intervals, which may vary from less than an hour to as long as a month or
more, are selected on the basis of the incubation or latency period of
the disease and the length of the period over which the cases are distri-
buted. In an epidemic of a disease having an incubation period measured
in hours (as with some foodborne diseases) with cases confined to a few
days, intervals of one hour or several hours would likely be used. With
diseases having an incubation period measured in days, daily intervals
would be more appropriate.

An appropriate interval for graphing cases is important to the subsequent
interpretation of the resulting epidemic curve. The most important kind
of error that can be made in this regard is the selection of an interval
that is too long, such as graphing cases of staphylococcal intoxication
according to the week or month of onset of symptoms. Such intervals
obscure small differences in temporal distributions, including secondary
waves of cases resulting from person-to-person transmission, and prevent

8

the graph from being used for its two principal purposes. A useful guide
to the selection of an interval for graphing cases is to make the in-
terval between one-eighth and one-fourth as long as the incubation period
of the disease in question. It often is desirable to make several epidemic
curves, each based on a different graphing interval, to find that one which
best portrays the data.

Epidemic Curves of Common Source and Propagated Source Outbreaks

Epidemics are often referred to as being either common source (cases
resulting from exposure to the same, common, source) or propagated source
(person-to-person transmission). In epidemics of some diseases both
types of sources might be involved, the initial cases resulting from
exposure to a common source, and subsequent (secondary) cases resulting
from person-to-person spread, as in figure 2. Some of the cases shown,
especially those which occured after July 8, might not have been re-
lated to the epidemic at all; they might be part of the endemic pattern
of occurrence of that disease.

Figure 2

Cases of Salmonellosis by Date of Onset
of Illness, Clarksville, Tennessee,
July 4 - 15, 1970
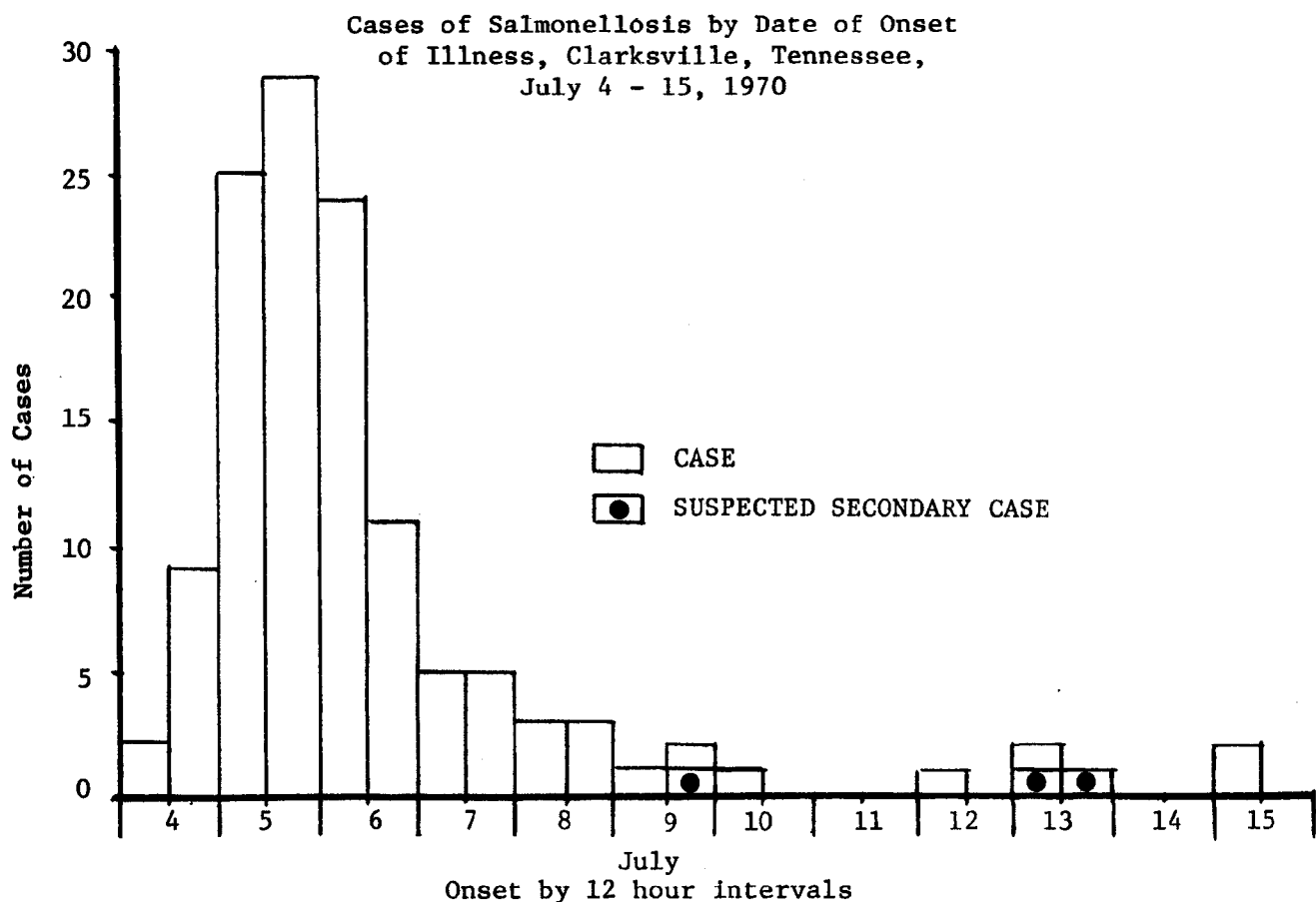
Source: *Morbidity and Mortality Weekly Report, USDHEW, PHS, HSMHA, CDC,
Vol. 19, No. 30, August 1, 1970, p. 296.*

The duration of an epidemic is influenced by:

1. the number of susceptible persons who are exposed to a source of infection and become infected;
2. the period of time over which susceptible persons are exposed to the source; and,
3. the minimum and maximum incubation periods for the disease.

Epidemics involving a large number of cases, with opportunity for exposure limited to a day or less, of a disease having a maximum incubation period of a few days or less, usually have an epidemic curve which approximates a "normal" distribution (Figures 3 and 4). When such epidemic curves are observed in epidemiologic practice, one can usually conclude that a common source was involved and that exposure of the cases to the source occurred over a short period (relative to the maximum incubation period of the disease).

Figure 3

Cases of Staphylococcal Intoxication by Incubation Period, Nashville, Tennessee, May 25, 1969
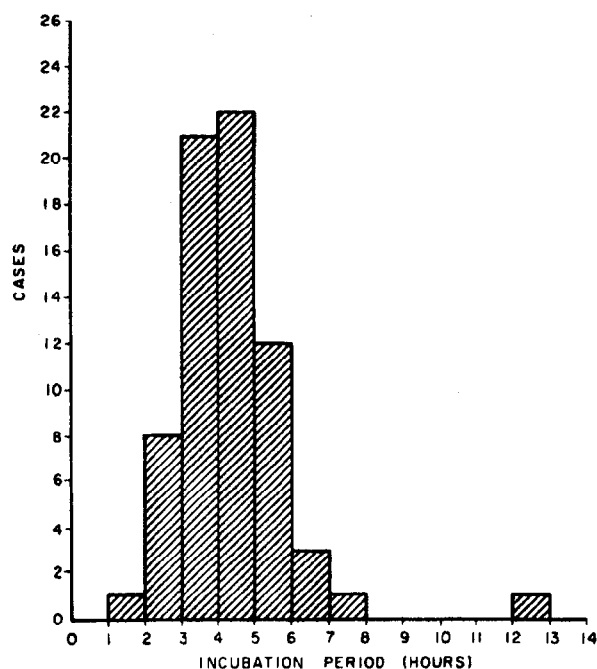
Source: Adapted from Morbidity and Mortality Weekly Report, USDHEW, PHS, HSMHA, CDC, Vol. 18, No. 34, August 23, 1969, p. 295

On the basis of the difference between the maximum and minimum incubation periods, the usual duration of an outbreak of this disease resulting from a single, brief exposure is 5 hours (6 hours - 1 hour).

The actual duration of the above outbreak was 7 hours.

Figure 4

Cases of Acute Gastroenteritis Due to Vibrio parahaemolyticus, by Incubation Period, Covington, Louisiana, August, 1972

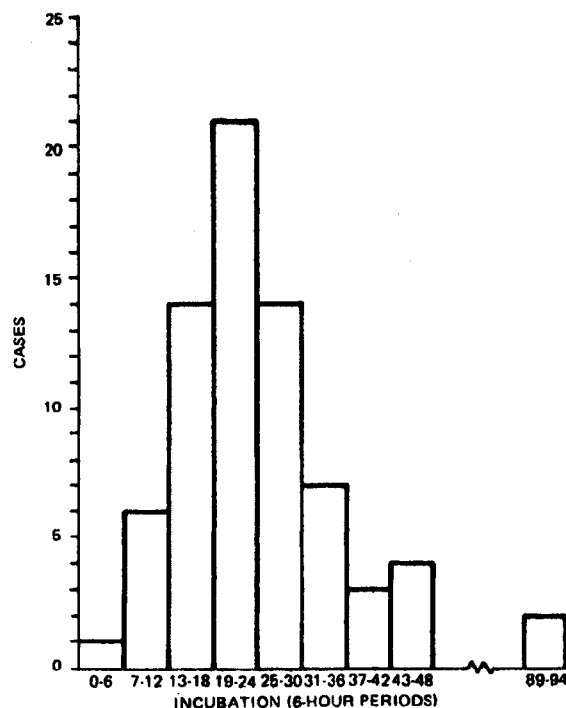Source: Adapted from Morbidity and Mortality Weekly Report, USDHEW, PHS, HSMHA, CDC, Vol. 21, No. 40, October 7, 1972, p. 342

On the basis of the difference between the maximum and the minimum incubation periods, the usual duration of an outbreak of this disease resulting from a single, brief, exposure is 46 hours (48 hours - 2 hours)

The actual duration of the above outbreak was 48 hours.

11

Exposure to a common source over a period of days, weeks, or months can be continuous or intermittent. With such prolonged exposure to common or propagated sources the epidemics lengthen considerably, as shown in figure 5. Intermittent exposure to a common source would produce a curve having irregularly spaced peaks.

Figure 5

Measles Cases By Week of Onset, Des Moines County, Iowa--July 31, 1971-Jan. 15, 1972
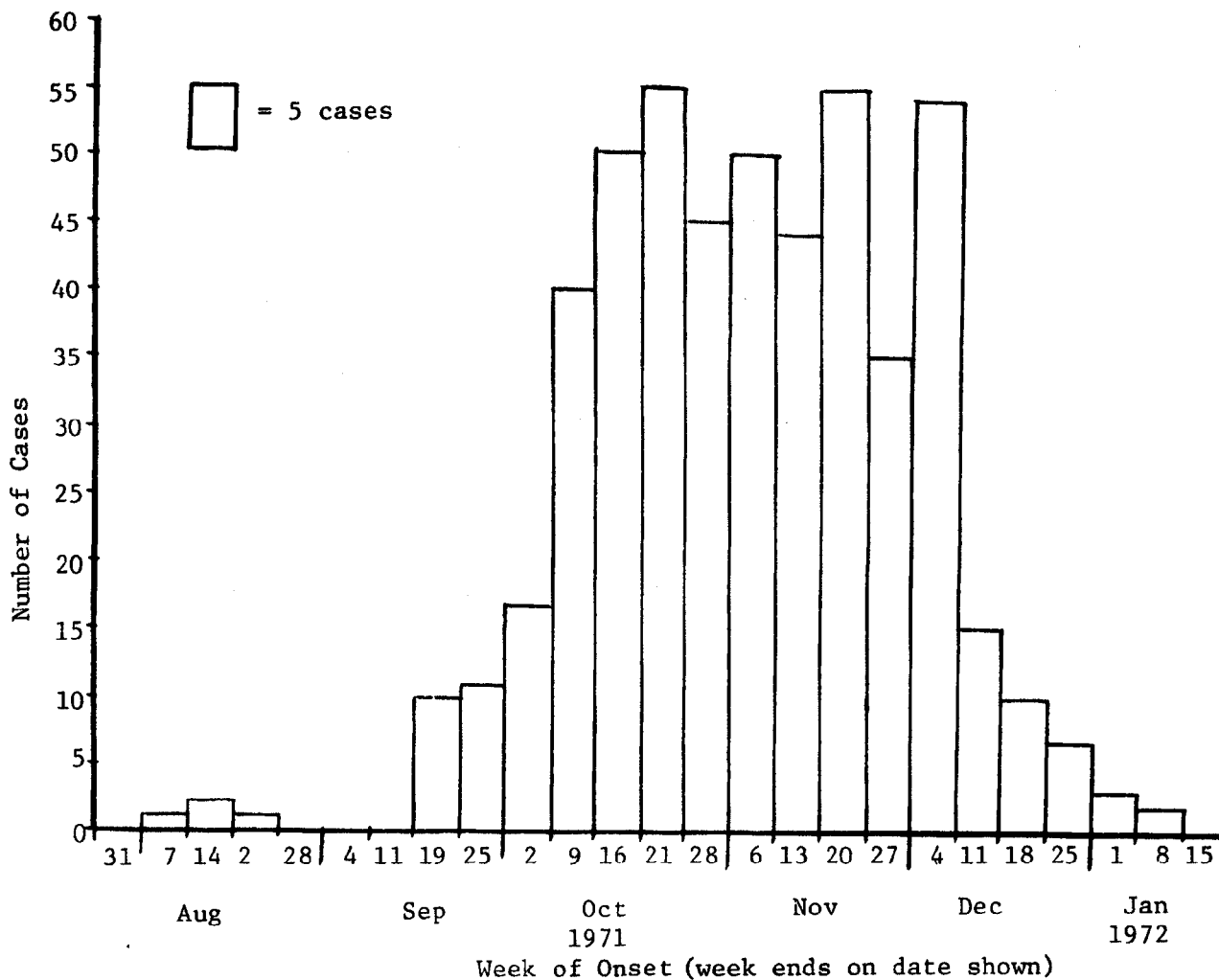
Week of Onset (week ends on date shown)

In propagated source epidemics the cases occur over a longer period than in common source epidemics of the same disease. Here again, though, the length of the incubation period influences the duration of propagated source epidemics. The cases shown in figure 6 resulted from person-to-person spread. This epidemic of viral hepatitis reveals a distribution of cases over several months.

Figure 6
Infectious Hepatitis Cases, By Week of Onset
Barren County Kentucky, June 1970 through April 1971

Week of Onset of Illness
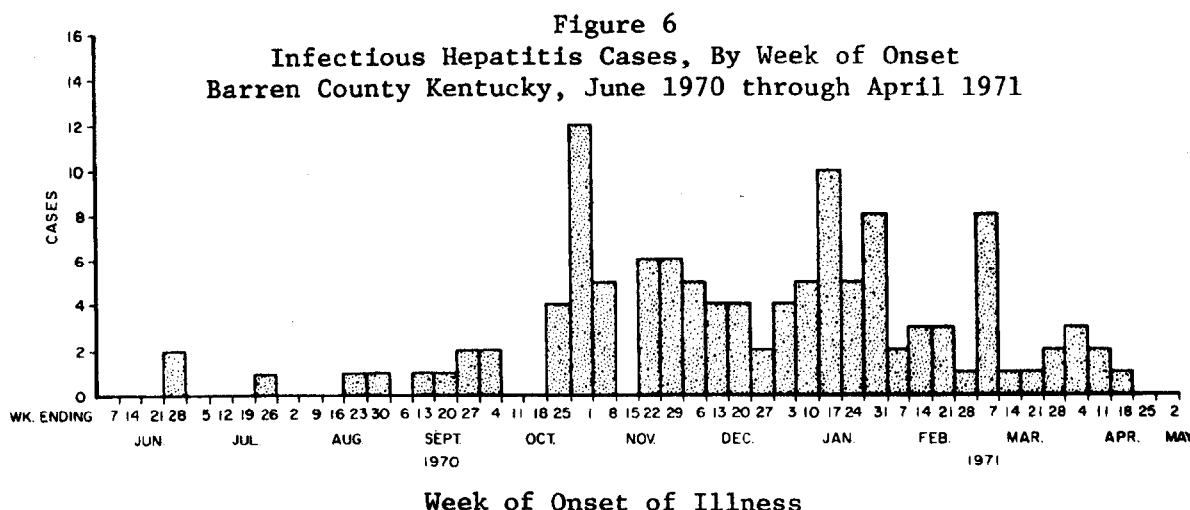
Source: *Morbidity and Mortality Weekly Report, USDHEW, PHS, HSMHA, CDC,*
Vol. 20, No. 15, April 17, 1971, p. 137

Explosive epidemics due to person-to-person transmission are less commonly seen. When they do occur, a disease having a short incubation period is usually involved. If secondary and tertiary generations (figure 7) occur, the intervals between peaks often approximate the average incubation period of the disease.

Figure 7
Measles Cases, By Date of Onset
Aberdeen, South Dakota, October 15, 1970 through January 16, 1971

Day of Onset of Illness
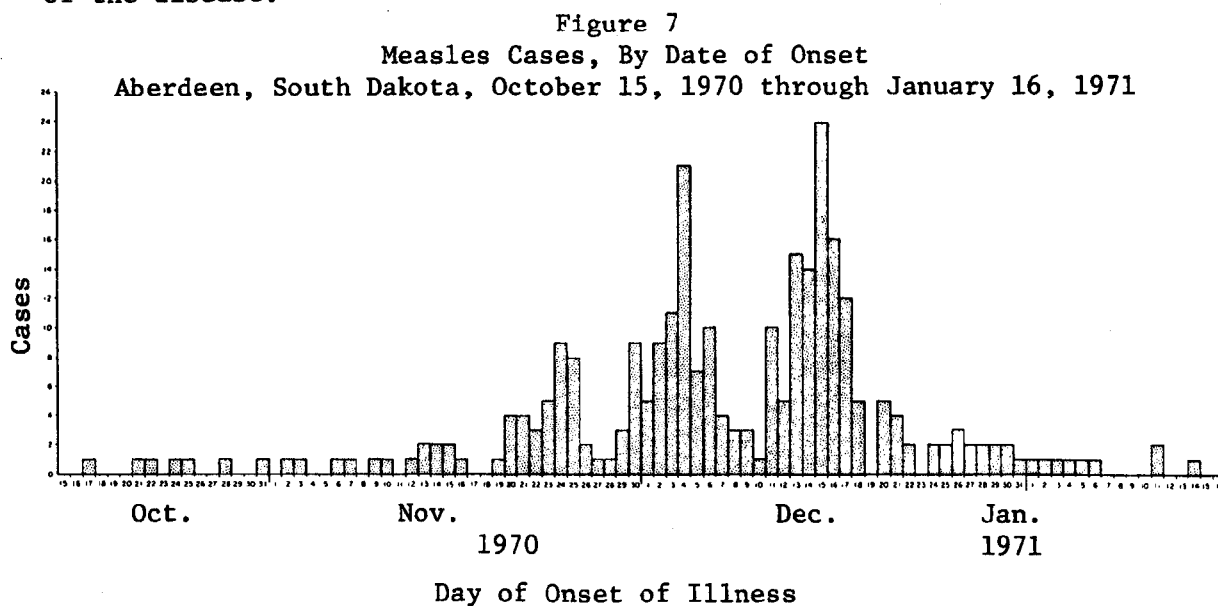
Source: *Adapted from Morbidity and Mortality Weekly Report, USDHEW, PHS,*
HSMHA, CDC, Vol. 20, No. 4, January 30, 1971, p. 26

Determination of the Probable Period of Exposure of Cases in a Common Source Outbreak

Knowing the mean, maximum and minimum incubation periods of a specific disease being investigated and the dates of onset of the cases, the most likely time of exposure of the cases to the source can be identified. There are two common methods of doing this. The first uses the mean incubation period (which can be obtained from Control of Communicable Diseases in Man). To use this method, it is necessary to identify the date of the peak of the epidemic or the date of the median case and count back in time one incubation period. For example, referring to the rubella outbreak in Figure 8, since the peak occurs on the 25th and since the mean incubation period is about 18 days, the probable time of exposure of the cases to the source is found to be on or about June 7th.

Figure 8

37 Cases of Rubella, Sun City, June 21-29

June
Date of Onset
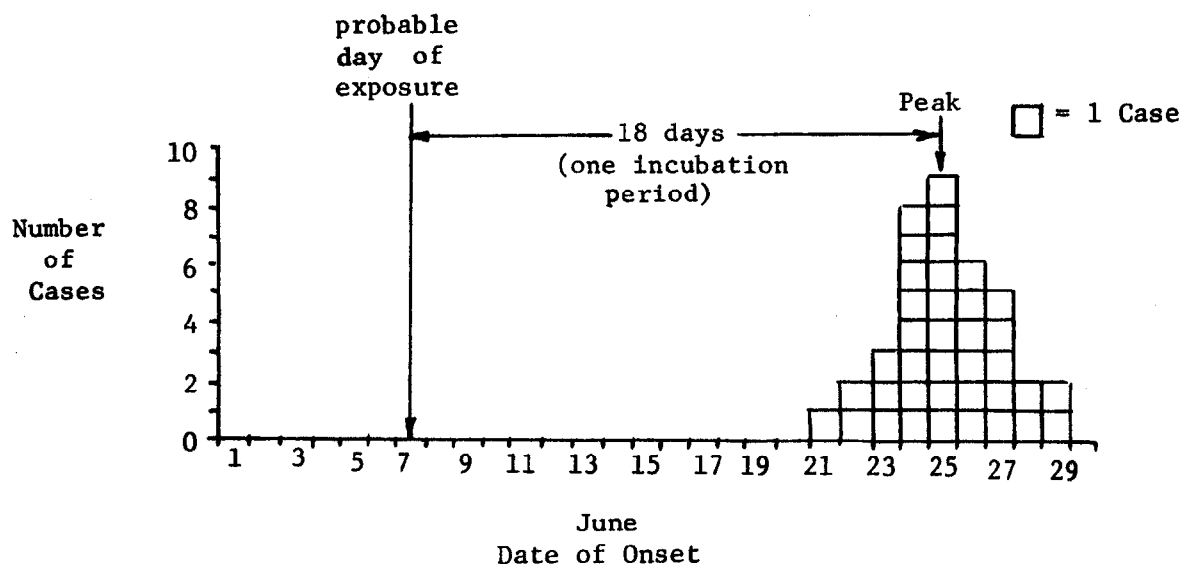
The other method is to use the minimum incubation period and count back in time from the first case and use the maximum incubation period and count back in time from the last case. Referring once again to our rubella example the minimum and maximum incubation periods are approximately 14 and 21 days respectively, and the probable period of exposure by this method (Figure 9) is therefore from June 7 to June 8.

The resulting interval in both methods may be lengthened on either side of the calculated dates of probable exposure to ensure inclusion of the actual period of exposure.

Figure 9

37 Cases of Rubella, Sun City, June 21-29
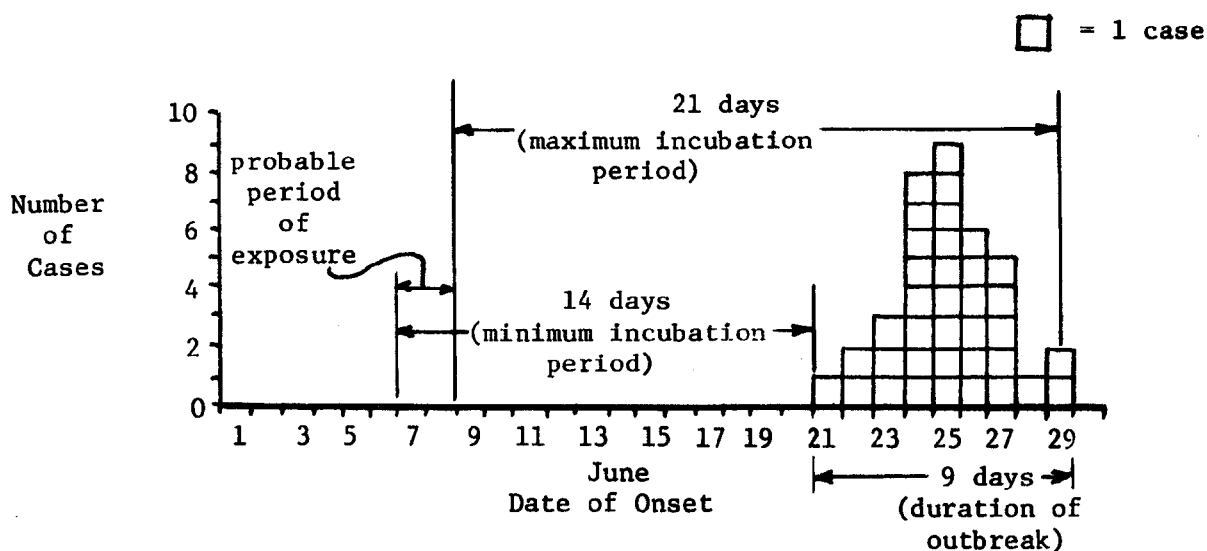
These methods can only be applied, though, when the duration of the epidemic (9 days in Figure 9) is approximately the same as or less than the difference between the maximum and minimum incubation periods for the disease in question (21 days minus 14 days, or 7 days, in Figure 9). If the duration of the epidemic is significantly longer than this interval, then the epidemic might be due to a continuing common source or a propagated source or both.

Figure 10 is an example of an outbreak of infectious hepatitis subsequent to a brief exposure (one day) to a common source of infection (fruit punch served at a luau). Other outbreaks of infectious hepatitis in which the period of exposure of the cases to the source was limited to one day or less have been typified by distributions similar to this. With exposure of one day or less and given that the incubation period is from 15 to 50 days, one would expect the duration of the subsequent epidemic to be no longer than 35 days (50 minus 15). That the duration of the epidemic (24 days) was less than expected more strongly supports the conclusion of a brief duration of exposure.

Figure 10

Cases of Hepatitis A in Individuals Attending a Luau, By Day of
Onset of Illness, Orange County, California, 1971
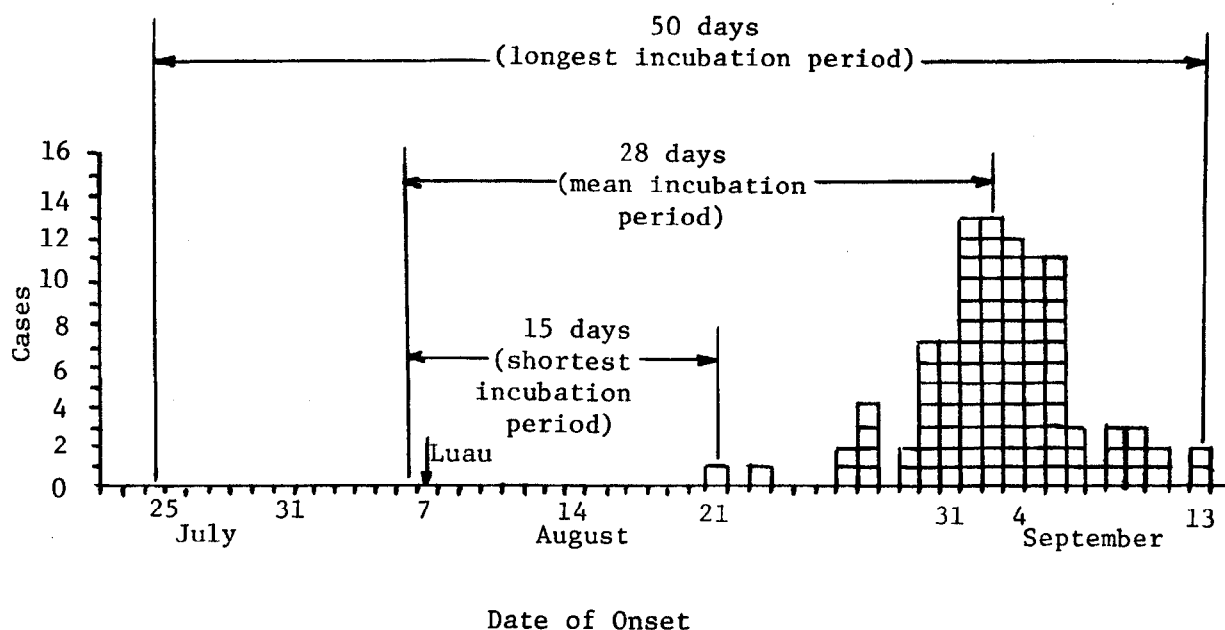
Date of Onset
*Date of Onset for 1 Case Undetermined

*Source: Hepatitis Surveillance, USDHEW, PHS, HSMHA, Report, No. 35,
July 1972, p. 12*

This example shows two limitations of the minimum/maximum method of identifying the probable period of exposure. First, counting back 15 days from the first case produces a date of August 6, the day before the actual date of exposure (rather than, ideally, producing either the actual date of exposure or a date one or two days after exposure). This could have been the result of any of several causes: (1) the first case is not really hepatitis, (2) it is hepatitis but the case was exposed somewhere else-- and before the luau--, (3) the case had an atypically short incubation period, or (4) the date of onset is incorrect.

The second limitation is that counting back 50 days from the last case produces a date of July 25, which is twelve days before the exposure. The result is an excessively long suspected period of exposure. This situation is a consequence of the outbreak lasting only 24 days, which is eleven days less than its theoretical maximum. In this case then, and generally, the probable period of exposure usually is more accurately identified using the mean incubation period.

Figure 11 shows another example of an outbreak of hepatitis A resulting from exposure to a common source of infection (drinking water at a school). Although in this case the exact number of days of exposure could not be established, it probably did not exceed 5 to 7 days on the basis of the duration of the outbreak (which was about 28 days). In this example, both methods of identifying the suspected period of exposure yield a date or a period that is consistent with the period of no chlorination of the school's water supply.

Figure 11

Cases of Hepatitis A, By Date of Onset,*
Colbert County, Alabama, October–November 1972
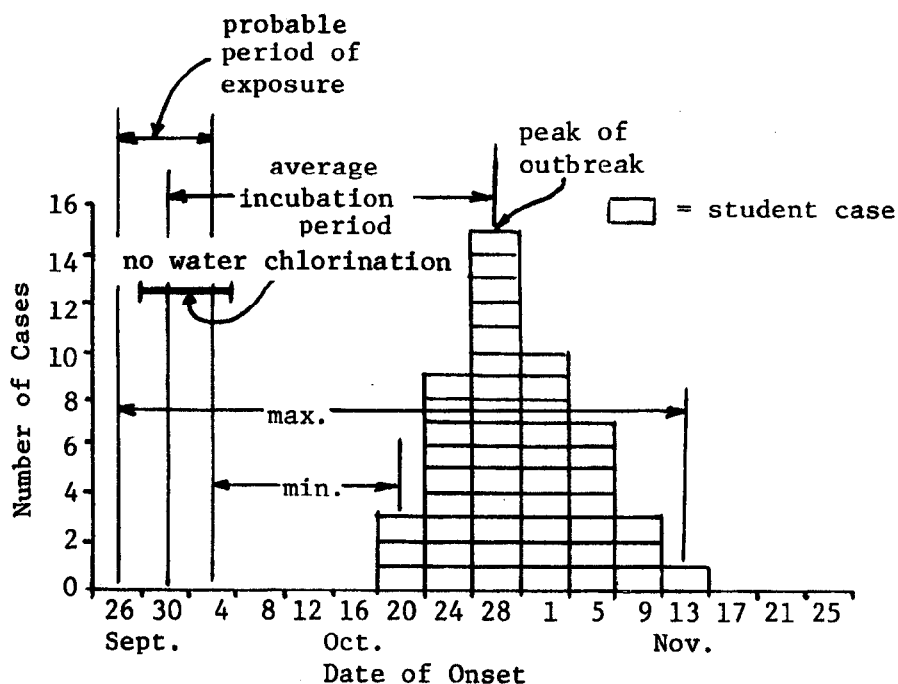
*Date of onset for two cases unknown

Source: Morbidity and Mortality Weekly Report, USDHEW, PHS, HSMHA, CDC, Vol. 21, No. 51, December 23, 1972

Identification of Secondary Spread Cases

To identify secondary cases of disease (e.g., among family members) first plot the onsets of illness of cases in time. Then, for successive cases in the same family, compare the intervals between cases to the length of the incubation period plus the period of infectiousness of the cases before their onsets. The following example (Table 2) is intended to illustrate this method. For this hypothetical disease, it is assumed that the illness and period of infectiousness last but one day. The incubation period for this disease is 2 to 5 days. An "x" indicates the day of onset of illness of each case in the family.

Table 2

Cases of a Disease ("X") that Occurred in Each
of Three Families, by Day of Onset

Family No.	Cases, by Day of Onset (August)									
	1	2	3	4	5	6	7	8	9	10
1		x		x				x		
2					x			x	x	
3			x							x

In the first family the interval between the time of onset of illness in the first and the second cases, and between the second and third cases, is consistent with secondary spread. In family 2, the second and third cases could both be secondary to the first case. That is, the interval between onsets is as great as or greater than the minimum and equal to or less than the maximum incubation period. In family 3 the interval between the first and second cases is greater than the incubation period of the disease and is not indicative of secondary spread from the first case.

Similarily, in family 2, the case occurring on the 9th would not be considered as secondary to the case occurring on the 8th since too little time lapsed between their onsets; and in family 1, the case occurring on the 8th is secondary to that occurring on the 4th, not the case occurring on the first.

The outbreak below shows the date of onset of jaundice in four cases of hepatitis A in a single family.

Case #	Date of Onset of Jaundice
1	January 12
2	January 15
3	March 3
4	April 14

The method just discussed can be used to identify which of cases 2, 3, and 4 could have been infected by case #1. For hepatitis A, the period of infectiousness can begin approximately three weeks before the onset of jaundice and last until three or four days after.

Figure 12

Dates of Onset of Jaundice in Four Cases of Hepatitis A,
and the Approximate Periods of Maximum Infectiousness and
of Possible Exposure of Each Case to the Source,
December 1976 through April 1977.

19
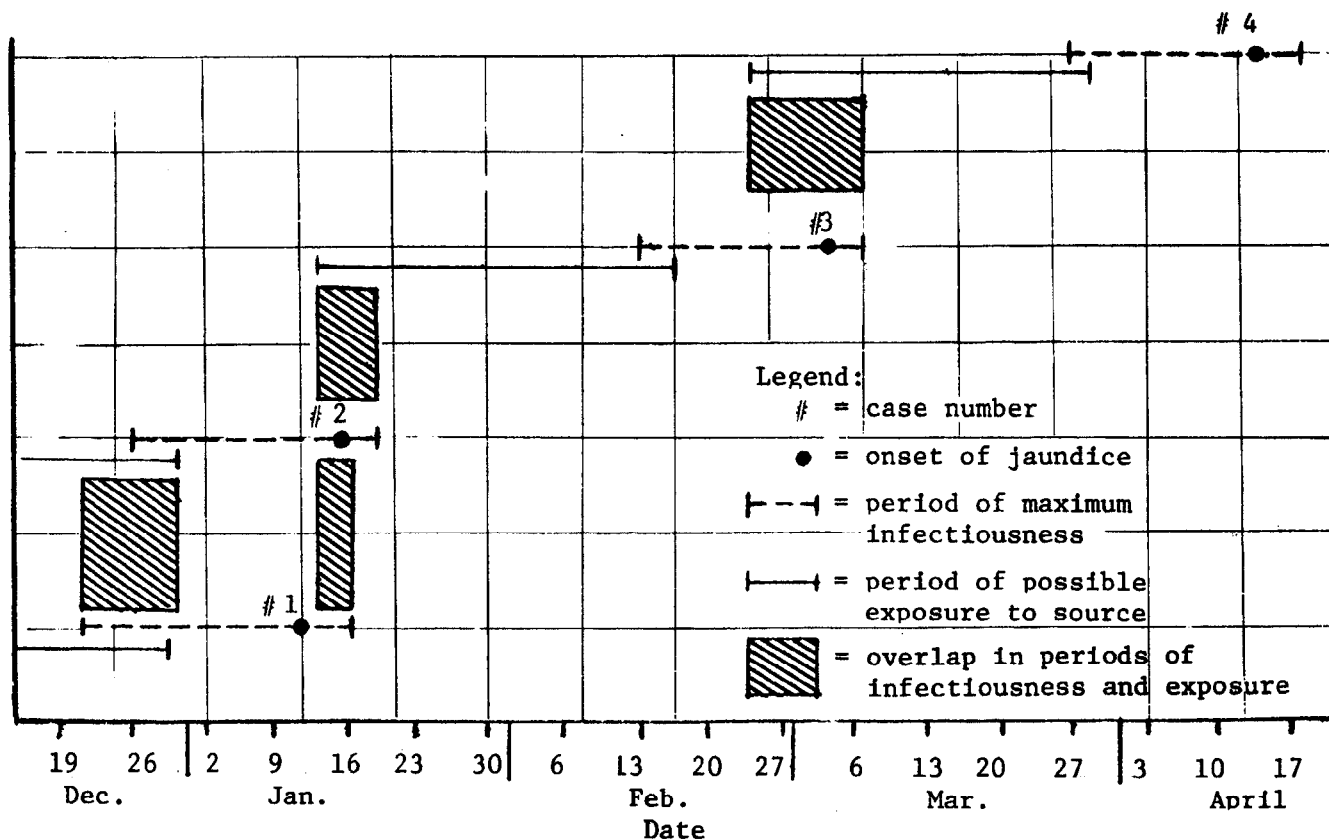
The next step is to identify those cases whose possible period of exposure overlapped another case's period of infectiousness. Doing this, one can observe that cases two and three could have acquired their infection from case #1. Case #1 could have been infectious during the period December 22 - January 16. Given that the incubation period ranges from 15 to 50 days, case #2 could have been infected by case #1 during the period December 22 through 31 and still develop symptoms by January 15. Similarly, case #3 could have been exposed either to case #1 on January 14 or 15 or to case #2 between January 14 and 18 and not develop symptoms until March 5 (50 days later, which would be the maximum incubation period). Case #4 could have been infected only by case #3 of these known cases.

In general, the characterization of an epidemic by variables of _time_ will be considered as adequately done when:

1. The time interval by which the cases are graphed is adequate to allow an identification of the probable period of exposure.
2. All known cases have been graphed, by the date of the onset of their symptoms.
3. The curve has been identified as typifying either a common source or a propagated source outbreak or as a combination of the 2 types.
4. In the instance of a common source outbreak, the following dates or periods have been identified:
    a. The peak of the outbreak
    b. The beginning, end, and duration of the outbreak
    c. The probable period of exposure of the cases to the source.
5. Further, if the source is both common and propagated, the known or suspected propagated cases have been identified and shown as such on the graph.

Place

The information gathered during the case count should provide clues to the population at risk according to place. This, combined with other information, should aid in the identification of the source of infection and the mode of transmission.

Spot maps of cases (Figure 13) are made to identify any patterns in case distributions according to place. Having the addresses of the cases and a map of the area, dots or pins representing the cases are placed on the map to show their distribution by place of residence. Clustering of cases, which may appear in conjunction with certain geographic areas such as census tracts, sanitary districts, and school areas, should be looked for. If clustering does occur, associations with possible sources of infection--such as water, milk, or food supplies--may become apparent.

20

Figure 13

Geographic Distribution of Measles Cases* by Place of Residence,
Waterbury, Connecticut; Jan. 31--April 4, 1971

*Two cases from southwest Waterbury are not shown.

Source: MMWR 5-8-71, Page 165

If clustering by place of residence is not apparent, it may be that the
appropriate "place" was not used. For example, mapping human cases of
brucellosis by place of residence may turn out to be unrevealing where-
as mapping them by place of occupation may provide the necessary clue to
the source. It may also be that despite the lack of apparent clustering,
the spatial distribution is still significant. If the disease agent was
spread by the airborne route, then the observed pattern might be explained
by the direction of the winds at the time of exposure of the cases to the
agent. If the disease agent was waterborne, then the wide geographic
dispersion of cases might mean that the entire population is or was at
risk of exposure.

Whatever geographic pattern is observed on spot maps, however, an assessment
of geographic variations in the risk of exposure or of infection must take
into account the distribution of the population. That is, area-specific
attack rates must be calculated, and conclusions about differences in risk
in different areas must be based on the rates and not on cases alone.
This is illustrated in table 3. Notice that Chicago, the area having the
highest number of cases, had one of the lowest area-specific attack rates.
The reverse is true of Evergreen Park.

Table 3

Attack Rates per 100,000 population by Locality of Confirmed and
Probable Cases of SLE (Encephalitis or Aseptic Meningitis),
Chicago SMSA, 1975

PLACE OF RESIDENCE	NO. CASES	POPULATION	ATTACK RATE
Chicago	90	3,366,957	2.7
Oaklawn	8	60,305	13.3
Evergreen Park	10	25,487	39.2
Blue Island	3	22,958	13.1
Des Plaines	2	57,239	3.5
Balance of Cook County	68	1,959,423	3.5
DuPage County	11	491,882	2.2
Will County	20	249,498	8.0
McHenry County	1	111,555	0.9
Kane County	2	251,005	0.8
Lake County	0	382,638	0.0
Unknown	5	-	-
TOTAL SMSA	220	6,978,947	3.2

It sometimes is productive to organize and analyze cases according to places
they have visited or traveled to or through (as in table 4).

Table 4

Diarrhea Attack Rates, By Water Service,
Community "A" and Community "B", August, 1975

Water Service	Number of persons			Attack rate (%)
	Ill	Well	Total	
Community "A"	98	57	155	63.2
Community "B"				
Not exposed to Com. "A" water	9	132	141	6.4
Visited Community "A":				
Drank water	22	18	40	55.0
Didn't drink water	0	6	6	0
Total Community "B" water	31	156	187	16.6

In the table above it can be seen that the attack rate for residents of
Community "B" is far lower than that for Community "A." However, when
the cases in Community "B" are tabulated according to whether they had
visited and drank water in Community "A," the attack rate in those who had
was found to be similar to the rate among residents of Community "A."

22

Other special situations occur with regards to cases in institutions of various kinds. For example, if the cases were hospital employees or patients, they should be analyzed according to their respective areas of work or stay: floor, ward, room, service, or bed. If investigation indicated an association with a school, the "place" information might be organized and analyzed according to classrooms within the affected school. (Table 5).

Table 5

Cases of Measles, and Attack Rates, by Grade,
Ganado Public School, Ganado, Arizona, April 1976

Grade	Cases	Population	Attack Rate (%)
K	24	85	28.2
1	17	86	19.8
2	7	61	11.5
3	8	90	8.9
4	4	104	3.8
5	23	99	23.2
6	12	95	12.6
Spec. Ed.	5	12	41.7
Total	100	632	15.8

An example of how the risk of illness may vary not only by the place of occupation, but by the specific time at which a person was at work is shown in table 6.

Table 6

Association Between Illness (Hepatitis A) in
Restaurant Employees and Working at the Restaurant
on the Evening of November 15 or 16, 1975

Worked on Evening	Number of Persons			
of November 15-16	Ill	Well	Total	%Ill
Yes	10	12	22	45
No	2	26	28	7
All Employees	12	38	50	24

The analysis of an epidemic by place will be considered to be adequately performed when the incidence rates for the constituent subareas reveal that the population in one or more of these subareas is at a significantly higher than average risk of exposure.

23

People can be described in terms of their inherent or acquired characteristics
(such as age, sex, race, immune status, and marital status), their activities
(form of work, play, religious practices, customs, etc.), the circumstances
under which they live (social, economic, and environmental conditions), and
perhaps other ways as well. These characteristics, activities, and conditions
are important since they determine, to a large degree, who is at the greatest
risk of acquiring specified infections or of experiencing other undesirable
health conditions.

As in the analysis by time and place, associations among cases may be evident
at an early stage and enable one to focus upon one or more of the characteris-
tics, activities or conditions which appear above. Analysis of the cases by
age is frequently the most important, productive, procedure in the analysis of
person variables since age is, in general, more strongly related to disease
occurrence than is any other single characteristic of person. The tendencies
shown, for example, in tables 7 and 8, for persons in one or more age groups
to have a significantly higher attack rate than persons in other age groups
is not uncommon. Such patterns often provide clues which are invaluable in
formulating hypotheses regarding possible sources of infection.

Table 7
Attack Rates (per 100 persons), by Age Group,
of Cases of Diarrhea, in Day-Care Center A

Age	Number of Children in Regular Attendance	Number of Children with Diarrhea	Attack Rate (Percent with Diarrhea)
1	20	17	85
2	19	15	79
3	39	13	33
4	39	4	10
5	38	5	13
>6	18	1	6
Total	173	55	32

Table 8

Incidence Rates per 100,000 Population, by Age Group,
of Confirmed and Probable Cases of St. Louis Encephalitis
(Encephalitis or Aseptic Meningitis), in the Chicago
Standard Metropolitan Statistical Area, 1975

Age Group	No. of Cases	Population	Attack Rate
< 1- 9	8	1,299,952	0.6
10-19	15	1,333,796	1.1
20-29	21	1,014,357	2.1
30-39	20	808,917	2.5
40-49	23	858,176	3.5
50-59	35	757,321	4.6
60-69	49	512,255	9.6
70-79	32	286,632	11.2
> 80	17	107,811	15.8
Total	220	6,978,947	3.2

For analytic purposes, the incidence and distribution of cases by age is
often initially related to five-year age intervals. However, the investigator
should not automatically do this. If other age groupings permit the investi-
gator to make better inferences regarding the source of infection and the mode
of transmission, then those should be used instead. Other age groups commonly
used for various diseases are shown in table 9. Before using any set
of age groups, however, the investigator should be sure that accurate
denominator data (populations) can be obtained for the age groups preferred.

Table 9
Age-Groupings Commonly Used to Tabulate Age
Distributions of Cases of Selected Diseases

Diphtheria, Viral Hepatitis, Salmonellosis, Tetanus, and Meningococcal infections	Syphilis (P&S), Gonorrhea	Tuberculosis	Trichinosis, Leptospirosis	Measles, Rubella
Less than 1 year old				< 1 year
1-4		0-4 yrs.	0- 9 yrs.	1- 4
5- 9	0-14 yrs			5- 9
10-14 or 10-19		5-14	10-19	10-19
15-19	15-19			15-19
20-24 or 20-29	20-24	15-24		
25-29	25-29		20-29	
30-39	30-39	24-44	30-39	20+
40-49	40-49		40-49	
50-59	50+	45-64	50-59	
60+				
		65+		
			70+	
Total	Total	Total	Total	Total

In general, it is preferable to tabulate case data into relatively small
age groups, at least initially. These can be combined later into larger
groups if desired. The problem with the use of large age groupings is
that they may hide differences in the risk of illness which would be
valuable in pointing to possible sources of infection. For example, if
a school milk supply were contaminated and served as a source of infection,
the use of 5-year age groups would allow one to focus the investigation on
school-age children by revealing that persons in the pre- and post-school-
age populations were not ill and therefore presumably were not exposed.
If, instead of 5-year age groups, ten-year or greater age groupings were
used, such inferences would be difficult, if not impossible, to make.

Similar tabulations of cases according to other characteristics of persons usually must be done. A clue as to which of these might be of value often can be found among the characteristics of the cases. If certain characteristics are observed to recur among the cases (e.g., of one sex or the other) then categories of cases can be established (e.g., male and female). An example of this is shown in table 10 below. In the outbreak that this data relates to, it was relatively easy to determine early in the investigation that an assessment of the risk of illness by occupation would probably be useful.

Table 10

Job- and Sex-Specific Attack Rates for Epidemic Illness
in a Medical Care Facility, August-September 1975

Job	Number Ill	Total Population	Attack Rate (%)
Registered Nurse	13	197	6.6
(ICU Nurses)	(12)	(30)	(40.0)
Nursing Assistant	5	68	7.4
Ward Clerk	5	15	33.3
Licensed Vocational Nurse	4	41	9.8
Surgery Technician	1	3	33.3
Orderly	3	13	23.1
Laboratory Technician	3	41	7.3
Housekeeping Technician	1	37	2.7
Supply Worker	1	36	2.8
Home Care Coordinator	1	2	50.0
Physical Therapist	1	4	25.0
Engineer	1	8	12.5
Secretary	1	67	1.5
Other	0	164	0.0
Male Employees	5	136	3.7
Female Employees	35	560	6.3
All Employees	40	696	5.7

The characterization of cases by person is considered as satisfactorily performed when significantly different attack rates are found among persons having and persons not having one or more specified attributes; or that the attack rates vary significantly with the degree to which persons have the attribute.

Information Needed

Clearly, the kind of information needed to characterize an outbreak depends upon the disease involved, the causes of the outbreak, and the objectives of the investigation. However, a few general statements can be made about the information that usually should be obtained and recorded for each case. This includes:

1. Identifying information about the case and family: name, age, sex, address, and telephone number.
2. The presence or absence of specified signs and symptoms, including their severity, degree or frequency; the date and time of the onset of illness; the duration of illness; whether or not the case was hospitalized, and if so, where and when; and the attending physician's name.
3. Any laboratory tests performed and the results.
4. Epidemiologic data relating to:
   a. Exposure to possible sources during the probable period of exposure.
   b. Exposure of other susceptibles during the case's period of infectiousness: names, places and dates.
5. The investigator's name and the date of the interview.

If, in the conduct of the investigation, it becomes necessary for the investigator to collect, handle or ship microbiological specimens, certain practices should be followed in order for the laboratory to be of maximum assistance. These practices include the following:

- Selecting only the most appropriate specimens on the basis of the disease suspected.
- When collecting specimens for serological examination:
  . Collect paired specimens, the first as soon after onset of illness as possible, and the second from two to four weeks later.
  . Do not allow whole blood to be frozen.
- Labeling the specimens adequately, including the patient's name, the type of specimen and the date collected; and your name and address. Include a brief clinical history of the case, and the etiologic agent(s) suspected.
- Packing the specimen in such a way that the container will not break, spill, or leak.
- Protecting the viability of the organisms during shipment.

The characterization of an epidemic by time, place and person, collectively, is satisfactorily performed when the characterization provides a satisfactory basis for the formation of a hypothesis regarding the source of infection and mode of transmission which, when tested, is confirmed.

Step 4: Identifying The Source of The Etiologic Agent and its Mode of Transmission

Identifying the source and mode of transmission of the agent may require more than one application of the cycle of hypothesis formulation and testing. For our purposes, a hypothesis is a statement, the investigator's "best guess," using the available information, regarding the explanation of the occurrence of some event. In the context of an epidemic investigation, hypotheses generally are formulated around a suspected etiologic agent, the source of infection of the cases, the period of exposure of the cases to the source, the mode of transmission, and the population that has been exposed or is at risk of future exposure. Depending on the nature, amount and quality of information available to the investigator, the hypothesis might be addressed to any of the foregoing or to several of them simultaneously.

The purpose of a hypothesis is to provide a logical basis for planning and conducting the various investigations necessary to achieve the objectives of an epidemic or outbreak investigation. Therefore, the hypothesis must be stated in such a way that it can be tested, and the results of the test should provide a clear answer to the question of whether the hypothesis is correct.

To develop a hypothesis:

1. Identify the objective you are trying to reach (for example, confirm a diagnosis).
2. Identify the available information relevant to that objective. Continuing our example of confirming a diagnosis, this information includes the signs, symptoms, and laboratory finding of the reported cases, and the specific criteria established for a case.
3. Derive a logical conclusion from the available information and state it as a hypothesis (these persons suspected as having disease "x" do, in fact, have disease "x"). If more than one logical conclusion can be reached, the investigator might concurrently establish an additional hypothesis; however, it is generally preferable to establish and test one hypothesis at a time.

As another example, suppose that an investigation has proceeded to the point that the cases have been characterized according to selected variables of time, place and person, and the objective now is to identify the source of the etiologic agent and the mode (and vehicle and/or vector) of transmission. The development of a hypothesis requires making a comparison between the known, actual, distribution of cases and the distributions which could be predicted to occur for each of the various possible kinds of sources and modes of transmission. Your hypothesis would state that the source and mode responsible for the predicted distribution of cases that most closely approximates the known distribution is the true source and mode in the current outbreak.

29

The data presented in table 11 is the actual age-group distribution of cases of diarrhea in a community. Clearly, all age groups are significantly affected. Of the various possible kinds of sources of infection and modes of transmission that could produce such a distribution, that of a contaminated public water supply must rate high in priority for consideration. The hypothesis based on these data would state that the agent was waterborne and that the source of the contaminated water was the public water supply. Remember, though, that this is still just a hypothesis and not an established fact. However, if the situation were sufficiently serious--people were dying--this hypothesis is well enough supported to be the basis for recommending to the community the control measure of boiling their water until further notice.

Table 11

Attack Rates of Diarrhea by Age
Group, Community "A," August 1975

Age	No. Persons Ill	Population Total	Attack Rate (%)
0-4	11	27	40.7
5-9	20	41	48.8
10-14	35	58	60.3
15-19	41	66	62.1
20-29	40	66	60.6
30-39	45	66	68.2
40-49	42	74	56.8
50-59	56	96	58.3
60-69	51	82	62.2
70-79	23	36	63.9
80+	12	17	70.6
Unk.	3	5	—
Total	381	636	59.9

Having established your hypothesis, further information must be gathered in order to confirm or reject it and to exclude all other possible explanations. The hypothesis in our first example may be considered successfully tested and accepted as true (i.e., the suspects have disease "x") when the investigator has been able to demonstrate that:

1. The clinical, laboratory, and other criteria of a case of disease "x" have been applied to and met by each case.
2. No other disease could have met the criteria established for disease "x."

The hypothesis in our second example could be considered successfully tested and accepted as true (that is, the hypothesized source and mode of transmission are the true ones) when the investigator has demonstrated that:

1. A significant difference in attack rates exists between those persons exposed and those not exposed to the suspected source.
2. Either no other mode of transmission was common to the cases or that another mode could not account for the observed age and geographic distribution of the cases.

Regarding foodborne diseases, a hypothesis about the vehicle of infection based upon ingestion of contaminated foods is classically developed by comparing illness rates between persons that ate and persons that did not eat the suspected foods. Table 12, an "attack rate" table, is an illustration of the method employed. Examining the figures in the two columns labeled "Attack Rate," and comparing the attack rate for persons who ate each specified food with the rate for persons who did not eat that food, it can be seen that the greatest differences in rates are for the barbecued pork and barbecue sauce. These two food items, therefore, are the most likely vehicles of the disease agent. The hypothesis would be that either the barbecued pork or the barbecue sauce, or both, was the vehicle of infection. Recovery of the agent from either of these foods and from the ill persons would confirm the hypothesis bacteriologically.

Table 12

Food History Data from Persons Attending Picnic, Nashville, Tennessee, May 25, 1969

Food or Beverage	Group A Persons Who Ate Specified Food				Group B Persons Who Did Not Eat Specified Food			
	Ill	Not Ill	Total	Attack Rate (Percent)	Ill	Not Ill	Total	Attack Rate (Percent)
Barbecued pork	70	18	88	79.5	0	26	26	0
Barbecue Sauce	59	12	71	83.1	11	32	43	25.6
Coleslaw	48	32	80	60.0	22	12	34	64.7
Beans	60	39	99	60.6	10	5	15	66.7
Bread	59	38	97	60.8	11	6	17	64.7
Butter	15	14	29	51.7	55	30	85	64.7
Chicken	14	29	43	32.6	56	15	71	78.9
Ice Cream	50	30	80	62.5	20	14	34	58.8
Lemonade	12	9	21	57.1	58	35	93	62.4
Cola	53	34	87	60.9	17	10	27	63.0
Orange Drink	10	4	14	71.4	60	40	100	60.0
Coffee	12	6	18	66.7	58	38	96	60.4

In Snow's investigation of cholera in London, his initial hypothesis after studying the death rates in the districts served by the two water companies was that the epidemic was due to the ingestion of contaminated water supplied by the Southwark and Vauxhall Company. However, the initial analysis could not exclude the presence of factors other than the water supply which may have varied between these two areas and accounted for the observed differences in mortality rates. In order to test his hypothesis, he focused attention upon the large area supplied by both companies. Within this area, many adjacent homes were supplied by different companies, and both companies served consumers of essentially similar conditions with respect to housing, economic level, occupation, and age. By gathering information about the source of the water supplied to each home in the area in which a cholera death occurred, and grouping the data by water source, he was able to confirm his initial hypothesis, since the cholera mortality rates for each source agreed with the rates for those areas served exclusively by each water company (table 13).

Table 13

Mortality from Cholera in London, July 8 to August 26, 1854,
Related to The Source of Water in Three Groups of Water Supply Districts

Water Districts, by Source	Water Supply of Individual Houses	Population, 1851	Cholera Deaths	Cholera Death Rate per 1,000 Population
1. Southwark & Vauxhall Co.	S. & V. Co.	167,654	738	4.4
2. Lambeth Co.	Lambeth Co.	19,133	4	0.2
3. Both Co's.	S. & V. Co.	98,862	419	4.2
	Lambeth Co.	154,615	80	0.5
4. Rest of London	All Sources	1,921,972	1,422	0.7

It is not always possible to demonstrate that a hypothesis is correct. Failure to confirm a hypothesis might occur for several reasons: the hypothesis might be wrong; or it might be correct but was badly stated, the test was not valid or adequate or was badly done, or the needed evidence was not available. If the evidence indicates that the hypothesis is false then a new hypothesis must be formulated and tested. If the evidence is inconclusive then possible causes for this must be identified and eliminated.

The source of infection and mode (and specific vehicle or vector) of transmission can be considered to have been correctly identified when the respective hypothesis have been tested and found valid.

Step 5: Identifying Populations at an Elevated Risk of Infection

When the source and mode of transmission have been confirmed, persons at an elevated risk of exposure must be identified and appropriate preventive and control measures implemented. Exactly who is at an elevated risk of exposure depends on the agent, the nature of the source, how the agent is transmitted, and the various attributes of susceptible individuals that increase the likelihood of their being exposed to the agent. In table 14 are several hypothetical examples in which an etiologic agent, a source of infection and mode of transmission are specified for a series of outbreaks, and the resultant characteristics of the population at risk of exposure described.

Table 14

Examples of the Specification of the Primary Population at Risk
of Exposure to the Source of Infection as a Function of
the Agent, Source and Mode of Transmission

Agent	Source of Infection	Mode of Transmission	Primary Population at Risk of Exposure to the Source
1. Hepatitis A virus	Bakery	Foodborne; contaminated glazing on donuts	Susceptible individuals who eat glazed donuts prepared at the bakery during the period that the human source of contamination made them. Also, contacts of cases (a possible continuing common source)
2. Clostridium perfringens	Banquet at a hotel	Foodborne	Persons who attended the banquet. (a common source, but brief exposure)
3. Measles virus	A particular school	person-to-person	Any contacts with no history of either the disease or immunization; primarily persons attending the school, and extracurricular and home contacts (propagated source)
4. Shigella sonnei	A carrier in one ward of a mental institution	Contact with the carrier or the contaminated environment of the carrier	Other residents of the ward; employees who work in that ward; and visitors (a continuing common source)

5. Hepatitis B virus	Units of blood used in a particular hospital	Bloodborne	Patients that receive transfusions; and medical, nursing and other personnel in that hospital that administer the transfusions (a possible continuing common source).
6. St. Louis encephalitis virus	Infected mosquitoes in a particular geographic area	Mosquito-borne	Susceptible persons residing in areas where the mosquito is breeding and becoming infected; especially those whose environment includes much standing water, clean or polluted, in natural or artificial containers; and those whose home is not screened or whose activities take them out of doors at night in breeding areas.

Whether or not the population at risk has been completely identified will be known when either of two conditions occur: future cases arising from the source occur only among persons thought to be at high risk, or, preferably, control measures aimed specifically at these populations prevent the occurrence of future cases.

Implementing Control Measures

When the general characteristics of the high risk population have been described as in the preceding table it then becomes necessary to identify the specific preventive and control measures appropriate to the populations involved. The control measures subsequently implemented might be addressed to any or all of these (and others): the source of the cases' infection, the original source, the means by which the agent is transmitted, and the susceptibles who are at a high risk of exposure.

Selected control measures may be initiated as early as the medical diagnosis of a case. The administration of immune serum globulin to family contacts of a case of Hepatitis A is a good example. Other measures may be initiated at varying points in time, depending upon the disease involved, sources of infection and modes of transmission. If contaminated foods are responsible, they may be destroyed. If water should be found (or seriously suspected) to be a source of infection, utilization may be discontinued until the supply and distribution system are decontaminated or it may be continued with public warning to boil the water prior to consumption. If contact with a contaminated source is involved, then steps can be taken to prevent contact with the source until such time as it can be eliminated. Immunization, early diagnosis, and treatment represent other methods of control which may be employed as the situation dictates.

The rapid initiation of practical and efficient control measures is the most valuable means of measuring the successful completion of an epidemiologic investigation.

Reports of Epidemic Investigations

The primary purpose of the epidemic report is to enhance the likelihood that the experience gained and the discoveries made are put to the best possible use: the design and implementation of improved surveillance techniques and prevention and control measures. Following is a suggested format for an epidemiologic report.

1. An introduction, which describes the circumstances leading to the initiation of the investigation.

2. A background, which briefly describes the setting in which the problem exists, including geographic, political, economic, demographic, and historical aspects.

3. A description of the studies conducted, including the reason (i.e., the hypothesis being tested), method, and source of information, for each. Examples of topics to be dealt with in this section are case-finding, verification of the diagnosis, use of control groups and samples analyzed.

35

4. The <u>results of the studies</u>, which include facts only, and specifically exclude attempted explanations, editorial comments, discussion and opinions. The data presented may deal with the previous as well as the current experience of the community with this disease. Examples of the data to be presented in this section are tabulations of characteristics of the cases (age, sex, race, occupation, etc.) and calculated attack rates; times of onset of illness of the cases (including an epidemic curve); results of laboratory tests; and other evidence which points to a probable source of infection or which rules out a suspected or possible source.

5. An <u>analysis of the data and conclusions</u>, which is an interpretation of the data for the purpose of accepting a hypothesis, and ruling out others, concerning the infectious agent, source of infection, reservoir, mode of transmission (including vehicle or vector), and high-risk groups. It would be appropriate to compare here the epidemiologic features of the current outbreak with other outbreaks.

6. A description of the action taken (control measures). This should include a specification of the objectives of the action taken, a discussion of the methods employed (how, when, where, and by whom) to achieve each, and a description of the effectiveness and cost of the control measures. The latter should include, among whatever other measures are used, the number of cases that develop subsequent to one incubation period after implementation of control measures until the time when the incidence rate returns to the pre-epidemic levels. And the cost of the control measures should be expressed both in terms of dollars and man-hours of time, by profession.

7. <u>A description of any other important outcomes</u>, such as:

   a. the effect of the epidemic on the population - health, legal and economic consequences.

   b. the effects of control measures on:
      The population - immune status, way of life
      Reservoirs - abundance; distribution
      Vectors - abundance; distribution
      Other life

   c. discoveries of new infectious agents, reservoirs, modes of transmission (including vehicle or vector).

8. Recommended procedures for the improved surveillance and control of the disease in the future. This might include references to surveillance data sources used, the scope and quality of the data to be collected, the frequency of collection, the timeliness and methods used in collecting, consolidating, analyzing and distributing the data, and the responsibilities of specific individuals within the health organization to perform the necessary tasks.